Appendix 2: Multiple imputation to account for missing data

From the core wave 2 ELSA sample (n=8780), 20.7% of cases had one or more missing values for the variables of interest; 9.3% (n=820) for ethnicity, 5.3% (n=464) for health literacy, 5.1% (n=446) for wealth, 2.5% (n=221) for depressive symptoms and less than 2% (n<153) for remaining variables. Little's MCAR test was significant ($X^2(64)=875.3$, $p<0.001$) indicating that values were not missing at random. Missing values were imputed using the fully conditional specification method within SPSS version 18.0 (an iterative Markov chain Monte Carlo method), to create 20 imputed datasets. Variables included in the imputation model were health literacy (as four separate items), all covariates in the substantive analyses, as described in the methods section (age, sex, wealth, ethnicity, occupation, education, limiting illness, limited activities of daily living, depressive symptoms, heart disease, diabetes, hypertension, cancer, stroke, asthma, lung disease, smoking, physical activity, alcohol consumption and cognitive function), mortality and survival in months. Sight problems were cited as a reason for non-completion of the health literacy test at interview and therefore self-rated sight (excellent to poor, or blind (n=33), on a 6-point scale) was also included in the imputation model.

A comparison between complete cases and the pooled imputed data is shown in Table A1.

Table A1. Comparison table to show differences between respondents with complete data and imputed data, for variables with 1% imputed data or above

	No. of imputed values	Original data, complete cases only (n=6945), n (%)	Imputed data, pooled (n=8780) n, (%)
Non-white ethnicity	820	168 (2.4)	210 (2.4)
Health literacy	464		
High		4667 (67.2)	5745 (65.4)
Medium		1404 (20.2)	1790 (20.4)
Low		874 (12.6)	1245 (14.2)
Wealth quintile	446		
Lowest (poorest)		1343 (19.3)	1740 (19.8)
Highest (wealthiest)		1424 (20.5)	1699 (19.3)
Depressive symptoms	221	992 (14.3)	1405 (16.0)
Occupational group	152		
Managerial		2342 (33.7)	2838 (32.3)
Intermediate		1692 (24.4)	2108 (24.0)
Manual		2911 (41.9)	3834 (43.7)
No. of words recalled, mean (SD)	144	5.7 (1.7)	5.6 (1.8)
No. of animals listed, mean (SD)	131	20.1 (6.3)	19.6 (6.6)
Mod/vigorous activity ≥1/wk	95	5337 (76.8)	6469 (73.7)
Correct date and time	93	5563 (80.1)	6885 (78.4)

Mean survival time in months based on all cases within the imputed dataset was 57.2 (SD 10.9), compared to 57.8 (SD 10.1) for complete cases. To reduce the influence of pre-terminal cognitive decline, participants who died within 12 months of interview were excluded (n=108 for complete cases; 169 for imputed data). We subsequently re-ran the Cox proportional hazards regression models described in the main manuscript, comparing complete cases with the pooled imputation data. The results are summarised in Table A2. Consistent with previous analyses, medium health literacy predicted mortality after adjusting for age and sex, HR 1.23 (1.02 to 1.50) p=0.035, but was not a significant predictor after taking into account indicators of socioeconomic position, HR 1.18 (0.97 to 1.43), p=0.102. The relationship between low health literacy and mortality was significant after adjusting for demographics, socioeconomic indicators, health behaviours and baseline health. When cognitive function was included in the regression model based on imputed data the low health literacy mortality association was no longer statistically significant, HR 1.21 (0.97 to 1.49), p=0.086, but remained of borderline significance within complete cases.

Table A2. Association between low health literacy and all-cause mortality. Hazard ratios with corresponding 95% confidence intervals from multivariable Cox proportional regression models with high health literacy as the reference category. Based on participants surviving more than 12 months post interview

	Hazard Ratios (95% CI), by Health Literacy Score					
	Complete cases, n=6837		p value	Pooled imputed data, n=8611		p value
Crude hazard ratio	2.81	(2.28 to 3.46)	<0.001	3.00	(2.50 to 3.60)	<0.001
Model 1: adjusted for age and sex	1.72	(1.39 to 2.13)	<0.001	1.75	(1.46 to 2.12)	<0.001
Model 2: model 1 + socioeconomic status and ethnicity	1.56	(1.25 to 1.94)	<0.001	1.56	(1.29 to 1.90)	<0.001
Model 3: model 2 + baseline health status*	1.46	1.17 to 1.82)	0.001	1.41	(1.17 to 1.68)	<0.001
Model 4: model 3 + health behaviours**	1.41	(1.23 to 1.76)	0.003	1.38	(1.13 to 1.69)	0.002
Model 5: model 4 + cognitive function***	1.26	(1.001 to 1.58)	0.049	1.21	(0.97 to 1.49)	0.086

*Baseline health status included long-standing limiting illness, limited activities of daily living, depressive symptoms and self-reported doctor diagnosed: heart disease, diabetes, stroke, cancer, asthma and chronic lung disease; **Health behaviours included current smoking, moderate/vigorous exercise once or more per week and daily or almost daily alcohol consumption; ***Cognitive function included orientation in time, immediate word list recall and fluency of animal naming

Sensitivity analyses

We considered that by imputing health literacy values for *all* participants, including those who were too ill to complete an assessment, the imputation risked introducing bias by imputing falsely high health literacy values. We therefore also repeated the multiple imputation procedure for those with a valid health literacy score at wave 2 (n=8,316), and imputed missing values for baseline covariates and survival time. Based on those that survived 12 months, low health literacy was a significant predictor of mortality in all models, including adjustment for cognitive function (fully adjusted HR 1.27 (1.03 to 1.55), p=0.024). In a further multiple imputation, we recoded respondents missing a health literacy value at wave 2 as a ‘low’ score (based on the assumption that those who refused to take the test would have been unable to achieve more than 2 out of 4) and repeated multiple imputation for missing values, which generated very similar results (fully adjusted HR 1.28 (1.06 to 1.54), p=0.010).